

# SMU Data Science Review

---

Volume 3 | Number 2

Article 13

---

2020

## Time Series Analysis of Offshore Buoy Light Detection and Ranging (LIDAR) Windspeed Data

Aditya Garapati

*Southern Methodist University, [agarapati@mail.smu.edu](mailto:agarapati@mail.smu.edu)*

Charles J. Henderson

*Southern Methodist University, [chasehenderson@mail.smu.edu](mailto:chasehenderson@mail.smu.edu)*

Carl Walenciak

*Southern Methodist University, [cwalenciak@mail.smu.edu](mailto:cwalenciak@mail.smu.edu)*

Brian T. Waite

*Southern Methodist University, [bwaite@mail.smu.edu](mailto:bwaite@mail.smu.edu)*

Follow this and additional works at: <https://scholar.smu.edu/datasciencereview>



Part of the [Applied Statistics Commons](#), [Data Science Commons](#), and the [Longitudinal Data Analysis and Time Series Commons](#)

---

### Recommended Citation

Garapati, Aditya; Henderson, Charles J.; Walenciak, Carl; and Waite, Brian T. (2020) "Time Series Analysis of Offshore Buoy Light Detection and Ranging (LIDAR) Windspeed Data," *SMU Data Science Review*. Vol. 3 : No. 2 , Article 13.

Available at: <https://scholar.smu.edu/datasciencereview/vol3/iss2/13>

This Article is brought to you for free and open access by SMU Scholar. It has been accepted for inclusion in SMU Data Science Review by an authorized administrator of SMU Scholar. For more information, please visit <http://digitalrepository.smu.edu>.

# Time series Analysis of Offshore Buoy Light Detection and Ranging (LIDAR) Windspeed Data

Aditya Garapati<sup>1</sup>, Charles Henderson<sup>1</sup>, Carl Walenciak<sup>1</sup>, and Brian Waite<sup>1</sup>

Master of Science in Data Science, Southern Methodist University, Dallas TX 75275  
USA {agarapati,chasehenderson,cwalenciak, bwaite}@mail.smu.edu

**Abstract.** In this paper, modeling techniques for the forecasting of wind speed using historical values observed by Light Detection and Ranging (LIDAR) sensors in an offshore context are described. Both univariate time series and multivariate time series modeling techniques leveraging meteorological data collected simultaneously with the LIDAR data are evaluated for potential contributions to predictive ability. Accurate and timely ability to predict wind values is essential to the effective integration of wind power into existing power grid systems. It allows for both the management of rapid ramp-up / down of base production capacity due to highly variable wind power inputs and integration of wind power into regional and national energy trading markets. Modeling successfully indicates that Autoregressive Integrated Moving Average (ARIMA) models, given data histories of one day at one minute intervals, provide the most useful forecasts, even when compared to more advanced modeling techniques such as Long Short Term Memory (LSTM) neural networks. These findings demonstrate the continued utility of long-standing autoregressive techniques and their more rapid time to train as an advantage over more complex machine learning techniques. To drive the operational utility of the analysis for users familiar with the data set provided by Pacific Northwest National Labs, a prototype web-based wind data exploration dashboard is also provided to allow users to conduct "on the fly" Exploratory Data Analysis (EDA) and identify best fit models.

## 1 Introduction

At the end of 2019, the total installed wind power capacity in the United States was 105,583 megawatts (MW) and the potential wind power capacity of the United States (at an 80 meter (m) turbine hub height) and its territories exceeded 10,640,000 MW [17]. This makes wind the largest source of renewable energy in the country [2]. Successful integration of wind power resources into the national energy infrastructure, though, poses a number of challenges. These include the variability of wind energy availability including ramp events (rapid changes in demand or surplus of electrical production), induced wear and expanded emissions as a result of cycling output of traditional power production facilities (such as coal, natural gas, or nuclear), and increased difficulty in managing and balancing variable demand with a stochastic input source [5].

The ability to forecast available wind energy as represented by speed and other environmental parameters is a key component of any solution to help energy operators successfully manage and integrate wind power into the existing energy grid.

Ayodele et al. surveyed the technical challenges associated with wind power generation and integration in their 2012 paper where they summarize the primary issues as “power system security, power quality, and power system stability” [3]. Power system security in general describes the ability of a system to “withstand disturbances without causing a breakdown of the power system.” Here the authors describe the ability of power generation and storage to compensate for dips in wind energy availability. They also reference the previous work by Billington and Huang to model wind power capacity with time series techniques [4].

Although dated and primarily focused on probabilistic methods to determine the adequacy of generating capacity, Billington and Huang demonstrate the viability of using long term collection of wind speed values to perform univariate time series analysis to predict future wind speed values at a given location. Specifically they identify an ARMA(4,3) model that outputs realizations which are then used as inputs to wind speed simulations for subsequent modeling [4]. The data used to fit this ARMA model was trained over an 8 year period for general wind site power evaluation. Our approach varies in our interest in specific predictions on much shorter time scales, while exposing the insights of longer term historical wind values (if any).

One technique to obtain historic wind values at potential wind farm sites in offshore locations, is through the use of dedicated wind Light Detection and Ranging (LIDAR) sensors. The U.S. Department of Energy (DOE) Office of Energy Efficiency and Renewable Energy (EERE) with the support of Pacific Northwest National Lab (PNNL) is currently supporting the deployment of an offshore wind energy evaluation effort using 20,000-pound buoys, known as WindSentinel, to measure meteorological and oceanographic parameters related to wind energy capacity. These buoys make use of LIDAR and other instruments to measure wind speed and direction, air and sea temperature, local barometric pressure, relative humidity, wave height and period, water conductivity, sub-surface currents, and other values in one second intervals [16].

Both historical and current buoy data from multiple locations are available to support model training and evaluation. Historical data includes data collected between 2014 and 2017 in offshore locations near both Virginia and New Jersey. Active collection is ongoing off the coast of Massachusetts providing the opportunity to evaluate the model against real-world data that is regularly being updated. While at various points in this research we focused on current and historic data, we ultimately determined that the data were not sufficiently unique to warrant evaluation of one over the other. As a result, our focus (randomly) was on the New Jersey data, but all findings and techniques can be applied to all of the datasets.

Additional data sources available to interested researchers, but not used in this research include the similarly collected 10-minute interval LIDAR data made available by the Danish firm Orsted from three unique offshore wind farm and meteorological station locations, and the Ding 2019 text that contains a number of datasets provided as exemplars to explore the methodologies described in the text from both on and off-shore locations [12,6]. While potentially useful, these datasets were excluded from this analysis due to the lack of clear usage restrictions or lack of specificity surrounding the context in which the data were originally collected. The anonymized collection of some of the datasets made them less desirable from a contextual understanding perspective and the overseas collected data made comparisons to existing U.S. wind resource references less impactful.

## 2 Sensors

### 2.1 LIDAR Background

LIDAR is widely used as a mapping and survey technology from a wide variety of platforms (vehicles, aircraft, fixed point collection). The National Oceanic and Atmospheric Administration (NOAA), for example, has used the unique properties of infra-red and green laser LIDAR data to collect both topographic and bathymetric data remotely [15].

When used in the context of wind measurement, LIDAR is coupled with well understood Doppler shift calculations on the back-scatter reflections of aerosols – small particulates contained in all air masses with diameters measured in fractions of microns – to obtain wind speed and direction values at a number of different ranges near simultaneously [13]. When deployed for offshore wind measurement LIDAR has almost entirely replaced the deployment of traditional meteorological towers with anemometer and wind-vane sensors that are generally fixed in the measurements they can provide based on their installation height and location. This ease of deployment has resulted in significant cost savings and addressed safety concerns associated with offshore meteorological tower construction and operation [10].

### 2.2 Vindicator and WindCube Sensors

The first version of WindSentinel deployed during 2014-2017 used a Vindicator III LIDAR sensor to measure wind speeds at a variety of heights. The more recent ongoing survey in Massachusetts uses a WindCube sensor. Both sensors face challenges with noise induced by environmental concerns, the foremost being that the sensor is mounted on a gimble on a floating platform and is still subject to some pitch, roll, and yaw adjustments that impact the variance of its measurements [16].

Multiple approaches to address this are used by PNNL and DOE including a smoothing function applied to the 1 second interval data (1Hz data) or the chunking of data into 10 minute averages.

In this research, we seek to explore statistical time series and machine learning approaches to enable the forecasting of wind strength, as a representation of

the available power, at given wind turbine hub heights with a forecasting horizon of 10 minutes. These prediction windows are selected to allow models to support load balancing and reserve power management in the immediate near term (10 minutes to one hour) [6].

### 3 Wind Modeling

In his book, “Data Science for Wind Energy”, Prof. Ding identifies two primary approaches to wind forecasting. The first, Numerical Weather Prediction (NWP), is built on modeling of atmospheric data to produce outputs similar to the weather forecast used to predict daily weather or hurricane events. This model generally performs well on longer time scales, but requires intense computing power to execute [6]. A second approach uses smaller scale, local sensor data to make statistical predictions based on single time series, sample-based modeling [6].

Our approach first explores the techniques and modeling methods associated with the creation of time series based forecasting of wind at a specific site given historical wind values on specified intervals and with co-located sensor data of interest. We then expand our time series based forecast by incorporating features derived from data collected more traditionally in support of NWP modeling (such as temperature, wind speed, wind direction, barometric pressure, and similar values) and the development of machine learning approaches that may contribute to or compete with the statistical time series prediction.

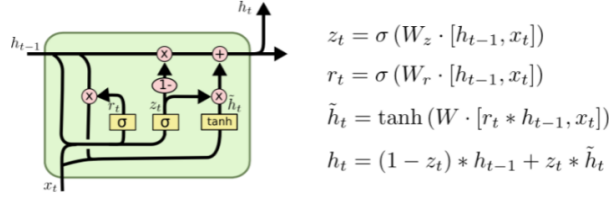
#### 3.1 Modeling Techniques

**Univariate Modeling Techniques: Autoregressive Integrated Moving Average (ARIMA) Models** ARIMA models represent a univariate methodology to model time series data by identifying relationships between historical values occurring at specific “lags” (in the autoregressive or AR component) and by modeling forecast errors due to white noise components (moving average or MA component). ARIMA models rely on the underlying data meeting the conditions of stationarity to provide predictions and can incorporate components to “model out” non-stationary components by transforming the data with differencing or seasonal terms.

**Neural Network Modeling Techniques: Long Short Term Memory (LSTM) Models** LSTM models are a variant of Recurrent Neural Networks and allow the neural network to retain memory across sequences of values and identify linkages to data that occur over longer time spans. This is ideal for time series analysis techniques where there may be correlation between inputs over time scales that may be lost by other recurrent or deep neural networks as they learn. Further, LSTMs overcome the vanishing gradient problem whereby weight updates in Recurrent Neural Networks become insignificant [14].

LSTMs help incorporate long term dependencies on the input vector by remembering information for a long period of time. The LSTM network architecture is a sequential model based upon two critical components, states and gates. The states include hidden state which depicts the value of previous hidden layer

and the input state which is a linear combination of current input data and hidden state. An illustration of a LSTM cell is shown below in Figure 1:



**Fig. 1.** An example LSTM Node and Possible Activation Function [9]

Fukuoka et al. demonstrate the use of a variety of neural network methodologies to perform time series forecasting over similar time horizons to the focus of our research. In addition to the use of historic wind values, the team also made use of solar radiation and wind direction as exogenous inputs to their models. Using a fully connected Deep Neural Network, LSTM, and a 1D-Convolutional Neural Network, they are able to obtain similar RMSE values of approximately 0.918 over a one hour prediction horizon for both models [8].

The LSTM was constructed by first creating a Deep Neural Network consisting of one input layer, three hidden layers, and a single output layer. The hidden layer nodes are then replaced with LSTM nodes that optimize using backpropagation and the squared error as the loss function [8].

The approach taken by Fukuoka et al. sets the underlying architecture that we derive our initial LSTM approach from and provides us with reference metrics that can be used to similarly evaluate our models, although we ultimately are able to only pursue a univariate LSTM approach [8].

## 4 Data Preparation

The PNNL BUOY LIDAR data is collected at 1 second intervals over a period of months and is therefore large, totaling over 42GB in size. Beyond the logistical challenges of data this size, the time scales we are exploring and the underlying processes we are attempting to describe are unlikely to be influenced significantly at this scale. Initial steps to prepare the data are therefore to determine an appropriate time scale and to average the values across that time horizon. For example, if time intervals of 10 minutes are desired, the wind speed values are binned into 10 minute intervals and then averaged. Missing wind speed values due to sensor malfunction or recording error are interpolated linearly with the previous and subsequent available records prior to averaging.

Meteorological data collected on the buoy was not natively collected at the same interval as the LIDAR sensor. Therefore additional cleaning and synchronization with the desired time interval was required. Collected at approximate

10 minute intervals, this data was either averaged across time spans greater than 10 minutes and matched to the appropriate period or again linearly interpolated for time scales shorter than 10 minutes requiring expansion of the “met” data. The non-instantaneous, fluid nature of wind and associated weather variables makes this sort of linear interpolation relatively reliable.

## 5 Exploratory Data Analysis

As an important step in understanding the distribution of the data used, exploration of the underlying distribution of the primary variable of interest was undertaken. Wind speed is highly variable and driven by a number of both local and regional processes. The distribution of wind speed is unique from other distributions as it is inherently non-negative and while low-speed events are relatively common, high speed events occur rarely and sporadically. A number of previous works identify the Weibull Distribution as being the most descriptive of Wind processes, but one of the more robust examinations of this distribution is that undertaken by Wais [18].

The Weibull distribution is a continuous probability distribution function that can be described using two or three parameters (the distinction of which when applied to wind problems being the focus of Weis’ work), but for the purposes here we will largely focus on the 2-parameter distribution. The function can be defined as:

$$f(t) = \frac{\beta t^{\beta-1}}{\eta^\beta} e^{-(\frac{t}{\eta})^\beta}$$

**Fig. 2.** Probability distribution function of 2-parameter Weibull distribution.

In this 2-parameter variant, the Scale and a Shape parameter are fit from the provided data using the Maximum Likelihood Estimation. The Ding text describes the shape parameter as affecting the skewness of the distribution and the scale parameter affecting its concentration or center [6]. Wherever applied in this analysis, the R “fitdistr” function was employed to fit a given data series to obtain the Weibull Scale and Shape parameters for transformation prior to standardization.

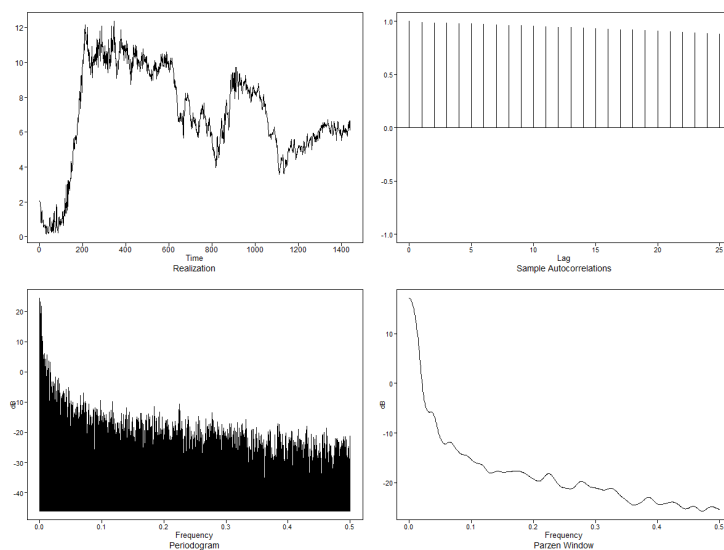
Referencing other research, Ding describes that data with a Weibull shape parameter of near 3.6 approximates the shape of a normal distribution. Therefore he recommends procedures to power transform the data by first fitting the parameters on the wind data and then solving for a power transform value “m” that is used to transform prior to further model fitting [6].

In his examination of fitting Weibull distributions to wind speed values, Wais identifies and ultimately determines that the two parameter Weibull distribution may not be an ideal fit for low-speed wind values. However, given the breadth of time contained in our data set and general confidence that site selection for the buoy deployment was made with the understanding that low wind values

would be relatively rare occurrences, we can move forward with confidence that the Weibull distribution accurately describes the wind values in our set.

After exploring these distributions, we found that transformation of the data according to the Weibull distribution did not have a significant impact on the ability of models to forecast the data. Weibull distributions, rather, may be useful if using probabilistic techniques to predict likely wind values according to a given distribution function. When used in linear models, on the other hand, the Weibull transform serves only as a scaling function on the data, but does not significantly improve (or harm) the linear relationships between the data and was ultimately determined to only introduce unnecessary complexity in the interpretation of results for minimal (if any) performance value.

### 5.1 Exploring a Representative Day



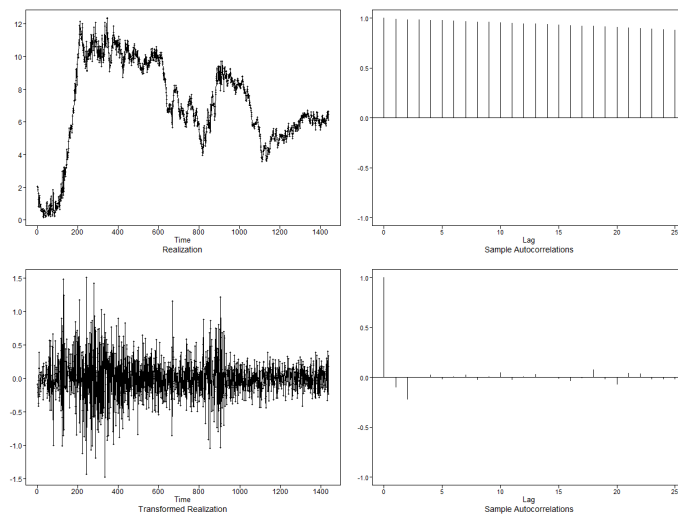
**Fig. 3.** Realization, Autocorrelation, Periodogram and Spectral Density Plot of Horizontal Wind Speed Values from 20 April 2016 at 1 Minute Intervals

As a first step to modeling over longer time horizons or larger datasets, the research team found it prudent to explore and fit a model for a single day as an initial exploration technique. Figure 3 shows a representative time series realization (upper left), the associated autocorrelations at various lags (upper right) and spectral density plots (lower right) which represent a Parzen Windowed smoothing of the Periodogram (lower left). As we can see, the realization demonstrates significant wandering behavior which is reflected in the spectral density with a strong peak at  $f = 0$ . The extremely slow damping behavior observed in the ACF



plot is an indication that the data is the result of a non-stationary process and should be differenced at least once before additional modeling.

In taking this first difference of the realization, we can see in Figure 4 that the autocorrelations and the differenced realization are much more conforming to our stationary conditions.



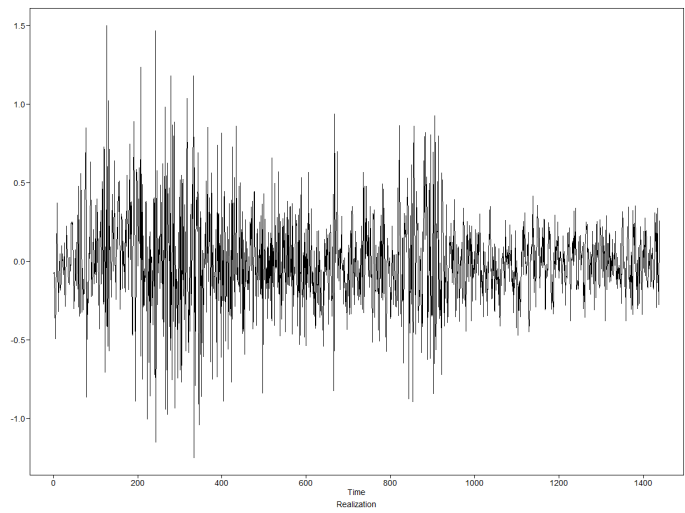
**Fig. 4.** April 20, 2016 horizontal wind speed data before and after first order differencing.

There does appear a change in variance at around time  $t = 1000$  that we would have to address if we were solely modeling this data alone. The residuals are indicative of an MA(2) component remaining after the transformation, but given that MA processes are known to not be good representations of real-world processes, we searched for other AIC-based model fit recommendations.

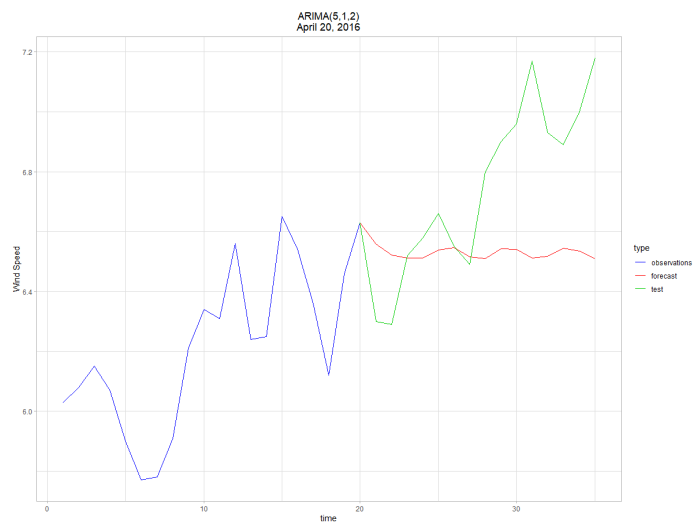
## 5.2 ARIMA Modeling

Using model selection parameters in the TSWGE R package [20] we selected an ARMA(5,2) to fit on this differenced data resulting in an overall ARIMA(5,1,2).

Modeling the differenced data provides residuals that are successfully converted to white noise as is confirmed by visual examination of the ACF plot in Figure 5. A Ljung-Box test performed on the residuals at  $K = 24$  and  $K = 48$  (p-values of .94 and .22 respectively) both fail to reject the null hypothesis that the residuals constitute white noise.



**Fig. 5.** Residuals of the 20 April, 2017 data after applying an ARIMA(5,1,2) model.

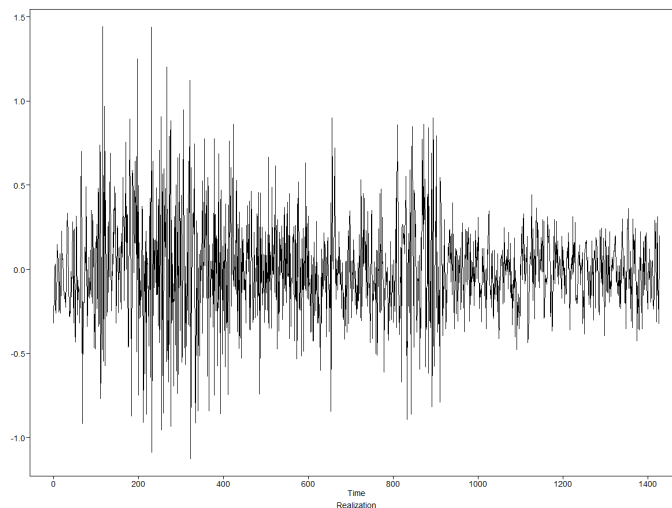


**Fig. 6.** Forecasts for 15 minutes ahead using fitted ARIMA(5,1,2) model on 20 April data.

As the  $t_0 + 15$  minute forecast using the ARIMA model in the figure above shows, the differenced term in the model dominates the behavior and forecasts are largely centered around the last observed value with some small cyclical behavior. When comparing these forecasts against the last 15 observed values, the forecasts achieve an Average Squared Error (ASE) of 0.1279<sup>1</sup>.

### 5.3 Vector Autoregressive Model (VAR)

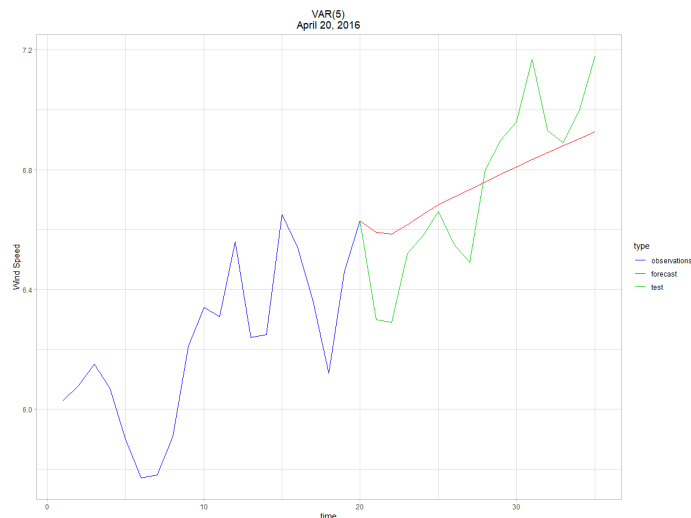
In an attempt to improve this initial model, we next attempted to fit a Vector Autoregressive Model (VAR model) to the data with additional exogenous variables consisting of horizontal wind direction, average air pressure and average humidity. A VAR selection procedure identified  $K = 5$  that provided optimal fit values by BIC. Looking at the residuals of this data we can see that the residuals are again observably whited. Ljung-Box confirms this at  $K = 24$  and  $K = 48$  (p-values of .79 and .21 respectively).



**Fig. 7.** Residuals after fitting the 20 April data with a VAR model and  $K = 12$ .

Forecasts using the VAR model as displayed above appear to provide a more satisfying fit. They capture the increasing trend of the wind values accurately and are not as dominated by the differencing term that caused the ARIMA model above to predict forecast values based primarily on the last observed value. This improvement is reinforced by the improved ASE of 0.03334.

<sup>1</sup> All references to ASE are measured by calculating error on the next 15 data points observed after the last time interval contained in the training set.



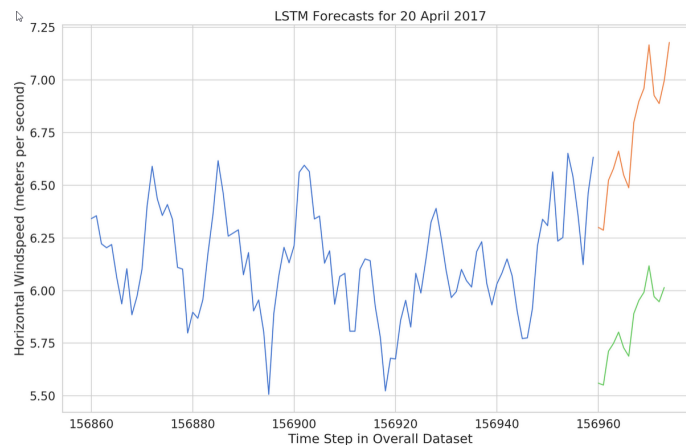
**Fig. 8.** Forecasts using the 20 April data with a VAR model and  $K = 12$ .

#### 5.4 Exploring LSTM models

Continuing the exploration of the single day data to familiarize ourselves with the modeling techniques and the potential that each model shows to accurately represent the wind data, we pursued a univariate LSTM approach. Even for univariate time series data, the configuration of LSTM models is non-trivial and can be at times difficult to interpret. Using the Tensorflow Keras API, we successfully configured an LSTM consisting of an input layer using 1442 samples (24 hours in 1 minute intervals), 3LSTM layers with 512 units each, a reduced layer of 50 units and final output layer of 15 units corresponding to the regression outputs for our 15 minute prediction interval. A dropout value of 0.2 was used between each LSTM layer and an Adam optimizer was used with default configuration. Similar to our autoregressive approaches, we used Mean Square Error as the primary loss function and terminated training to prevent overfitting after validation loss did not decrease beyond 0.001 for 20 epochs.

Multiple searches for additional hyperparameter tuning were conducted and included adjustments to the learning rate within the Adam optimizer, the number of nodes, modifications to both dropout and the recurrent dropout parameters within the LSTM itself. Ultimately the above configuration was found to be the most useful.

Despite these attempts at optimization, the LSTM model could only achieve an MSE of 0.636 on the same one day data and forecast window as the autoregressive models described above. It is interesting to note that the LSTM does appear to capture the trend of the data, but the ability of the model to learn different training sets varies significantly.



**Fig. 9.** Forecasts using the 20 April data with a univariate LSTM model.

Another perspective gained in this initial look at LSTMs is that the training complexity and time required to train these models has costs beyond the forecast predictive ability of the model that might be considered by anyone conducting the analysis. If time sensitivity and limited computational resources are to be considered, then the LSTM approach may not be useful.

## 6 Expanding Analysis to a Broader Dataset

Given the stochastic nature of the processes that generate wind as demonstrated by our exploratory analysis, where the ARIMA models were essentially attempting to model data at the edge of White Noise, training over large datasets (in this case with long histories of wind values) does not make sense or offer much to our forecasts. Consistent with our research into prior works and our own experience with the data, we settled on a one day training period to obtain 15 minutes ahead of forecasts for our final comparative analysis.

Once we made this determination, it became apparent that each model would need to be re-fit to the data for every 24 hour window being examined. This may be non-intuitive, but really is the best mechanism to ensure that the 15 minute forecast values are derived from recent, relevant and the most influential wind period. This is easy to understand when the reader considers that two days of no wind prior to an extremely windy day likely have no meaningful impact on the windspeed in any 15 minute period on the day being explored. As discussed briefly above, 24 hour "on demand" training was relatively easy to accomplish for the ARIMA approaches, but retraining the LSTM models for each 24 hour period can be lengthy and may make them less desirable in an operational context.

Because each 24 hour training window represents a unique realization of the wind speed data, average performance of the models over many iterations are not really representative of the model performance – because each model is uniquely generated for the time period it is fit. Rather, we chose to focus on the utility

of the models as they pertain to specific windows and selected a number of representative exemplars to include here to demonstrate the comparative utility of each model type. Including the 20th of April used for comparison above, we expanded our selection to other arbitrarily chosen dates to evaluate model performance on other 24 hour periods.

ARIMA				VAR				LSTM
Type	Train 24 Hr Period	15 Min Forecast	p	d	q	ase	p	ase
Base 1	2016-01-02	2016-01-03	BIC	BIC	BIC			
Base 2	2016-01-02	2016-01-03	AIC	AIC	AIC			0.41916
AIC	2016-01-02	2016-01-03	1	1	2	0.2611	11	0.153258
BIC	2016-01-02	2016-01-03	0	1	2	0.2617	4	0.092214
Base 1	2016-02-10	2016-02-11	0	1	2	0.9192		1.41737
Base 2	2016-02-10	2016-02-11	1	1	2	0.8983		
AIC	2016-02-10	2016-02-11	5	1	4	0.9515	15	0.8125
BIC	2016-02-10	2016-02-11	0	1	3	0.9033	4	1.17432
Base 1	2016-04-20	2016-04-21	BIC	BIC	BIC			0.63616
Base 2	2016-04-20	2016-04-21	1	1	2	0.1319		
AIC	2016-04-20	2016-04-21	5	1	2	0.1279	7	0.040312
BIC	2016-04-20	2016-04-21	0	1	2	0.1320	5	0.03334
Base 1	2016-07-02	2016-07-03	0	1	2	0.1770		0.23771
Base 2	2016-07-02	2016-07-03	BIC	BIC	BIC			
AIC	2016-07-02	2016-07-03	1	1	3	0.1618	19	0.293627
BIC	2016-07-02	2016-07-03	1	1	2	0.1789	5	0.221952
Base 1	2016-10-02	2016-10-03	0	1	2	0.099666		0.41507
Base 2	2016-10-02	2016-10-03	BIC	BIC	BIC			
AIC	2016-10-02	2016-10-03	4	1	4	0.12186	14	0.251624
BIC	2016-10-02	2016-10-03	1	1	2	0.126685	5	0.281799

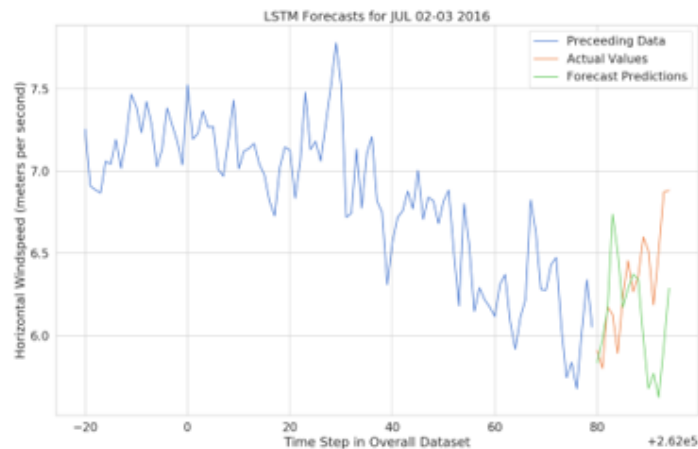
**Table 1.** Table of model Average Square Error results based on tested parameters and according to model type.

Table 1 above demonstrates the variation in results from these models. The models “Base 1” and “Base 2” represent fixed models that were best fit on an initial dataset and represent a sort of “control” group for model fit across multiple groups. In this case Base 1 is an ARIMA(0,1,2) and Base 2 is an ARIMA(1,1,2). As the table shows, in some instances, the recommended best fit according to AIC and BIC align with the fixed models and in others they do not, indicating that re-training continuously does provide some benefit.

The table also shows significant variation in the errors of each 24 hour period, perhaps reflecting the variability or complexity in each realization.

Figure 10 below shows the power of the LSTM when it is able to train a reasonable fit. The model, which achieved an ASE of 0.2377 in Table 1, is able to meaningfully capture both the trend and the fluctuations in the wind speed to

a meaningful degree. However, the ASE does not approach the average accuracy of the ARIMA models over the same period.



**Fig. 10.** Forecasts using the July LSTM data with a univariate LSTM model.

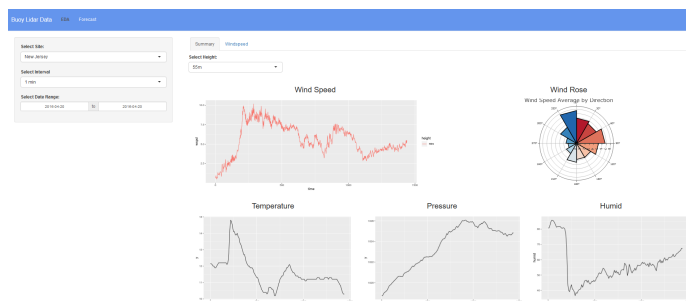
Finally, we can also observe that in general the ARIMA based approaches provide the smallest Average Square Error, but not always so. For example, the LSTM model appears to perform significantly better on the February 10th dataset, whereas the ARIMA models perform exceptionally on the October 2nd realization.

## 7 Operationalizing This Analysis

While this study does have at its core an academic objective, the team also sought to explore methods for making this time series analysis, the buoy lidar data, and forecasts derived available to an operational user seeking to make informed decisions. To this end, the team created an interactive operational “dashboard”<sup>2</sup> that offers the ability to conduct exploratory data analysis, to set modeling parameters, and to view the forecasts and error rates associated with each model type in “real time.”

In the EDA view, a user is offered the opportunity to configure the dataset selection, the time period of interest, and the “smoothing interval,” and altitude of measurement (height) to be used when displaying the data. They are presented with plots of the realization and associated exogenous variables. Realizations at different heights can be displayed in order to enable the user to discover potential relationships between data collected at various elevations.

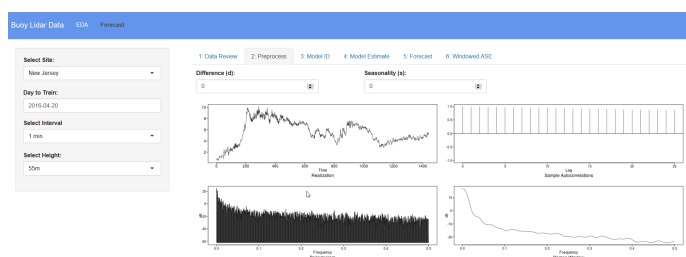
<sup>2</sup> At the time of this writing the dashboard is hosted online at <https://cwalenciak.shinyapps.io/windapp/> and is available for user interaction.



**Fig. 11.** Exemplar of the dashboard display for EDA use.

Transitioning to the forecast interface within the dashboard, the user is given the tools to evaluate specific realizations of the wind speed data, to apply pre-processing steps to meet stationarity assumptions and to identify potential best fit models. Once model selection has been completed, the forecasting dashboard allows the selected model to be fit with selected parameters and provides forecasts, errors, and windowed errors across selected forecast time horizons and datasets.

At this time, model fitting is restricted to the autoregressive and VAR modeling approaches enabled through TSWGE as the team used Python's interface to Tensorflow in order to build and train the LSTM models and has not had sufficient time to port the code to R. A similar interface to Tensorflow is available via an R API, but that integration into the dashboard is left for future work.



**Fig. 12.** Exemplar of the dashboard display for model fitting and forecasting.

We are excited to offer this tool to the wind prediction community and particularly those researchers who are working with the PNNL / DOE Buoy LIDAR datasets as we feel it offers a rapid means to interface with and evaluate the data in a lightweight fashion.



## 8 Ethical Considerations

### 8.1 Data Use and Protections

The Department of Energy is a cabinet-level agency that has missions in both energy and national security related matters. Its purpose is to implement policies that support national energy security and promote technological innovation in nuclear power, fossil fuels, and alternative energy sources. While much of the offshore wind energy LIDAR buoy data is associated with the Department of Energy, the data is buoy wind speed testing data that is openly available to the public and not considered a national security risk. Pacific Northwest National Labs, a Federally Funded Research and Development Center (FFRDC), cannot place any further restrictions on the data that were collected using public funds.

Department of Energy provides the buoy LIDAR data consistent with the guidelines associated with Open Data defined in the open data handbook [7]. This stipulates that data is available to all and for free use, as long as the original source of the data has been cited.

All data used was further reviewed by the research team to ensure protection of national security interests and avoid personal privacy information violations.

### 8.2 Potential Conflicts of Interest

No member of the research team has an identified conflict of interest associated with the data or the outcomes of the research. One member of the research team is a U.S. Government employee, but is not involved in energy policy, energy data analysis, or any matter related to the subject of this research. Another member of the research team is an employee of a national energy firm, but works in evaluating property tax compliance and cost evaluation and is similarly not involved in the wind energy or portfolio investment aspects of the company's operations.

### 8.3 Ethics of Wind Energy

It is beyond the scope of this project to evaluate in detail all the ethical risks of the promotion of wind energy. Generally, wind projects can interrupt flight paths and migratory patterns of some types of birds and bats leading to the deaths of many animals. Additionally, some individuals living near wind farms have surveyed negative feedback towards the aesthetics and noise caused by the turbines. With the adverse risks to wind energy, it is important to consider the placement of wind turbines in areas where the negative effects are minimized. Offshore wind turbines may limit these impacts, but pose other challenges to the communities and environment near where they are deployed.

Some community action to oppose offshore wind turbines as a blight on offshore views and due to other perceived environmental impacts have been documented [19, 11]. Wind turbines over the water can be moved to locations an acceptable distance away to lower noise and visual obstructions for humans. Offshore wind turbines can also be placed in locations with lower impacts to nature. Wind energy is considered a viable alternative to fossil fuels which is generally accepted to cause more harmful impacts to nature and global climate

change. Fossil fuels, when consumed in large quantities to generate power for cities and significant populations, can cause air pollution harmful to the health of people. Wind energy does not negatively impact air quality and therefore is an acceptable energy producing alternative.

## 9 Conclusion

Through this paper we have demonstrated that both statistical, autoregressive and machine learning LSTM models of wind speed data obtained from LIDAR sensors are feasible and provide benefit to the wind industry as they seek to predict and plan power resources in response to wind energy. However, we also identify that the stochastic nature of the processes that drive wind make any single approach to modeling wind speed difficult. Adding differencing terms to stationary models leaves very little to model in most cases and wind prediction is modeling components that are on the verge of being white noise. LSTM models attempting to do this in real time may be overly cumbersome to implement and train in any meaningful sense, particularly when more rapid ad hoc modeling can be performed using ARIMA techniques.

The continued value of wind modeling on these time scales may ironically be placed at risk by the deployment of horizontally situated “nacelle LIDAR” [1] to look forward into approaching wind masses at the site of wind turbines to reduce reliance on forecasting and obtain actual on-site values. We do believe, however, that modeling will continue to have a strong place in the wind industry as cost may prevent full implementation of horizontal LIDAR installations and these techniques may supplement the insights of those sensors or support energy management firms through periods of outage.

Overall, we are confident that our contribution will help to advance insights into the Buoy LIDAR dataset and that the provided dashboard will fuel both future interest and ease data exploration for similarly focused research.

## References

1. Zx tm turbine / nacelle lidar: Remote wind farm optimisation, <https://www.zxlidars.com/wind-lidars/zx-tm/>
2. American Wind Energy Association: Wind Energy Facts at a Glance (2020), <http://www.awea.org/Resources/Content.aspx?ItemNumber=5059>
3. Ayodele, T., Jimoh, A., Munda, J., Agee, J.: Challenges of grid integration of wind power on power system grid integrity: A review. *world* **3** (2020)
4. Billinton, R., Huang, D.: Wind power modelling and the determination of capacity credit in an electric power system. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* **224**(1), 1–9 (2010)
5. Bird, L., Milligan, M., Lew, D.: Integrating variable renewable energy: Challenges and solutions . <https://doi.org/10.2172/1097911>
6. Ding, Y.: *Data Science for Wind Energy*. Chapman and Hall/CRC (jun 2019). <https://doi.org/10.1201/9780429490972>
7. Foundation, O.K.: *Open data handbook*, <http://opendatahandbook.org/guide/en/>
8. Fukuoka, R., Suzuki, H., Kitajima, T., Kuwahara, A., Yasuno, T.: Wind speed prediction model using lstm and 1d-cnn. *Journal of Signal Processing* **22**(4), 207–210 (2018)

9. Github: Understanding LSTMs, <https://colah.github.io/posts/2015-08-Understanding-LSTMs>
10. Leosphere, Inc.: The role of lidar in offshore wind measurement. [https://www.lidar.gwu-group.de/files/2017\\_Leosphere\\_White\\_Paper\\_-The\\_role\\_of\\_Lidar\\_in\\_offshore\\_wind\\_measurement.pdf](https://www.lidar.gwu-group.de/files/2017_Leosphere_White_Paper_-The_role_of_Lidar_in_offshore_wind_measurement.pdf)
11. Miln-Tyte, A.: With more wind farms in the u.s. comes more debate (Jun 2019), <https://www.marketplace.org/2019/05/30/offshore-wind-energy-sparks-debate-on-long-island/>
12. Moore, D.J.: Offshore wind data., <https://orsted.com/en/Our-business/Offshore-wind/Wind-Data>
13. Pichugina, Y.L., Banta, R.M., Brewer, W.A., Sandberg, S.P., Hardesty, R.M.: Doppler Lidar-Based Wind-Profile Measurement System for Offshore Wind-Energy and Other Marine Boundary Layer Applications. *Journal of Applied Meteorology and Climatology* **51**(2), 327–349 (02 2012). <https://doi.org/10.1175/JAMC-D-11-040.1>, <https://doi.org/10.1175/JAMC-D-11-040.1>
14. Pulver, A., Lyu, S.: LSTM with Working Memory. arXiv e-prints arXiv:1605.01988 (May 2016)
15. U.S. Department of Commerce and National Oceanic and Atmospheric Administration: What is lidar (Oct 2012), <https://oceanservice.noaa.gov/facts/lidar.html>
16. U.S. Department of Energy: A2e: BUOY, <https://a2e.energy.gov/projects/buoy>
17. U.S. Department of Energy Office of Energy Efficiency and Renewable: WINDEXchange: U.S. Installed and Potential Wind Power Capacity and Generation (2020), <https://windexchange.energy.gov/maps-data/321>
18. Wais, P.: Two and three-parameter weibull distribution in available wind power analysis. *Renewable energy* **103**, 15–29 (2017)
19. Waters, M.: Anti-wind farm activism is sweeping europe-and the u.s. could be next (Oct 2018), <https://earth.gizmodo.com/anti-wind-farm-activism-is-sweeping-europe-and-the-us-c-1829627812>
20. Woodward, W.: TSWGE: Applied Timeseries Analysis, <https://cran.r-project.org/web/packages/tswge/tswge.pdf>